

This paper was presented at a colloquium entitled “Protecting Our Food Supply: The Value of Plant Genome Initiatives,” organized by Michael Freeling and Ronald L. Phillips, held June 2–5, 1997, sponsored by the National Academy of Sciences at the Arnold and Mabel Beckman Center in Irvine, CA.

Genome projects and gene pools: New germplasm for plant breeding?

MICHAEL LEE*

Department of Agronomy, Iowa State University, Ames, IA 50011

ABSTRACT Crop gene pools have adapted to and sustained the demands of agricultural systems for thousands of years. Yet, very little is known about their content, distribution, architecture, or circuitry. The presumably shallow elite gene pools often continue to yield genetic gains while the exotic pools remain mostly untapped, uncharacterized, and underutilized. The concept and content of a crop's gene pools are being changed by advancements in plant science and technology. In the first generation of plant genomics, DNA markers have refined some perceptions of genetic variation by providing a glimpse of a primary source, DNA polymorphism. The markers have provided new and more powerful ways of assessing genetic relationships, diversity, and merit by infusing genetic information for the first time in many scenarios or in a more comprehensive manner for others. As a result, crop gene pools may be supplemented through more rapid and directed methods from a greater variety of sources. Previously limited by the barriers of sexual reproduction, the native gene pools will soon be complemented by another gene pool (transgenes) and perhaps by other native exotic gene pools through comparative analyses of plants' biological repertoire. Plant genomics will be an important force of change for crop improvement. The plant science community and crop gene pools may be united and enriched as never before. Also, the genomes and gene pools, the products of evolution and crop domestication, will be reduced and subjected to the vagaries and potential divisiveness of intellectual property considerations. Let the gains begin.

The balance between food supply and demand will become more delicate as economies and human populations expand at various rates around the world. In most regions, crop production has satisfied or exceeded demand because of several factors and their interactions. The major factors typically have included natural resources (climate, arable land, crop germplasm, and a water supply), chemicals (fertilizers and pesticides), technologies (mechanization and plant breeding), and favorable socioeconomic conditions. There are indications that some sources (land, water, chemicals, and mechanization) are approaching the limits of their capacities to affect significant and sustainable gains in crop productivity in several regions of the world (1). Thus, gains in productivity will become more dependent on other sources.

In any scenario, plant breeding shall remain a vital component of effective agricultural systems. Plant breeding has contributed to increased crop productivity by systematically creating new genotypes (i.e., varieties) with superior adaptation to the needs of society, the resources of the production system, and the demands of nature in the target environments.

In the United States and probably elsewhere, the economic return on investments in plant breeding has been one of the great success stories in research and development (2). Worldwide, the fruits of plant breeding have stabilized economies and saved or enhanced billions of lives in the last 50 years.

Since the 1920s, when organized plant breeding programs were initiated on a large scale in the United States, there has been continuous improvement in the genetic component of productivity in major grain crops (3). Of course, those gains have been accompanied by some disturbing trends, such as decreased genetic diversity within the elite gene pool, increased genetic uniformity of the crop in production, and erosion of native genetic resources in the primary, secondary, and tertiary gene pools (4). The risks and consequences related to those trends remain at various levels for most crop species. For example, some suspect the primary gene pool of commercial U.S. maize has become sufficiently shallow such that the rate of genetic gain for grain yield may decline (5). This situation may be exacerbated in the short term as a result of consolidation within the seed industry and deployment of transgenic crop varieties. According to one assessment, the rate of gain for U. S. maize grain yields, from all sources, has declined by nearly 50% since the 1970s. However, the causes of the changes have not been determined (O. S. Smith, personal communication; Fig. 1). The same suspicion may be valid for other major crop species and granaries. Given the fundamental value and purpose of crop-based agriculture, it is essential that we enhance our appreciation of crop gene pools and our ability to improve them.

The first generation of genomic sciences produced several methods that improved our abilities to assess and manipulate a crop's gene pool by providing a new source of genetic information, DNA markers (e.g., restriction fragment length polymorphism, and simple sequence repeats). In many circumstances, marker-aided methods provided considerable advantages and power because, for the first time, data could be gathered and collated from many defined regions of crop genomes. Thus, DNA markers and comprehensive genetic maps have been used widely in assessing crop diversity and relationships (6, 7) and for developing more precise and rapid breeding schemes (8). Their use has replaced some methods and altered our perceptions of gene pools. Ultimately, these changes may foster development of deeper, broader, and richer gene pools capable of meeting the challenges of agriculture.

Assessments and Perceptions of Gene Pools. The genetic diversity and relationships of crop gene pools have been assessed by many methods and from several perspectives. The derived information has influenced the principles and practice of classification schemes, management and assembly of germplasm collections, risk assessments of crop genetic vulnerabil-

U.S. Average Maize Yields since 1920

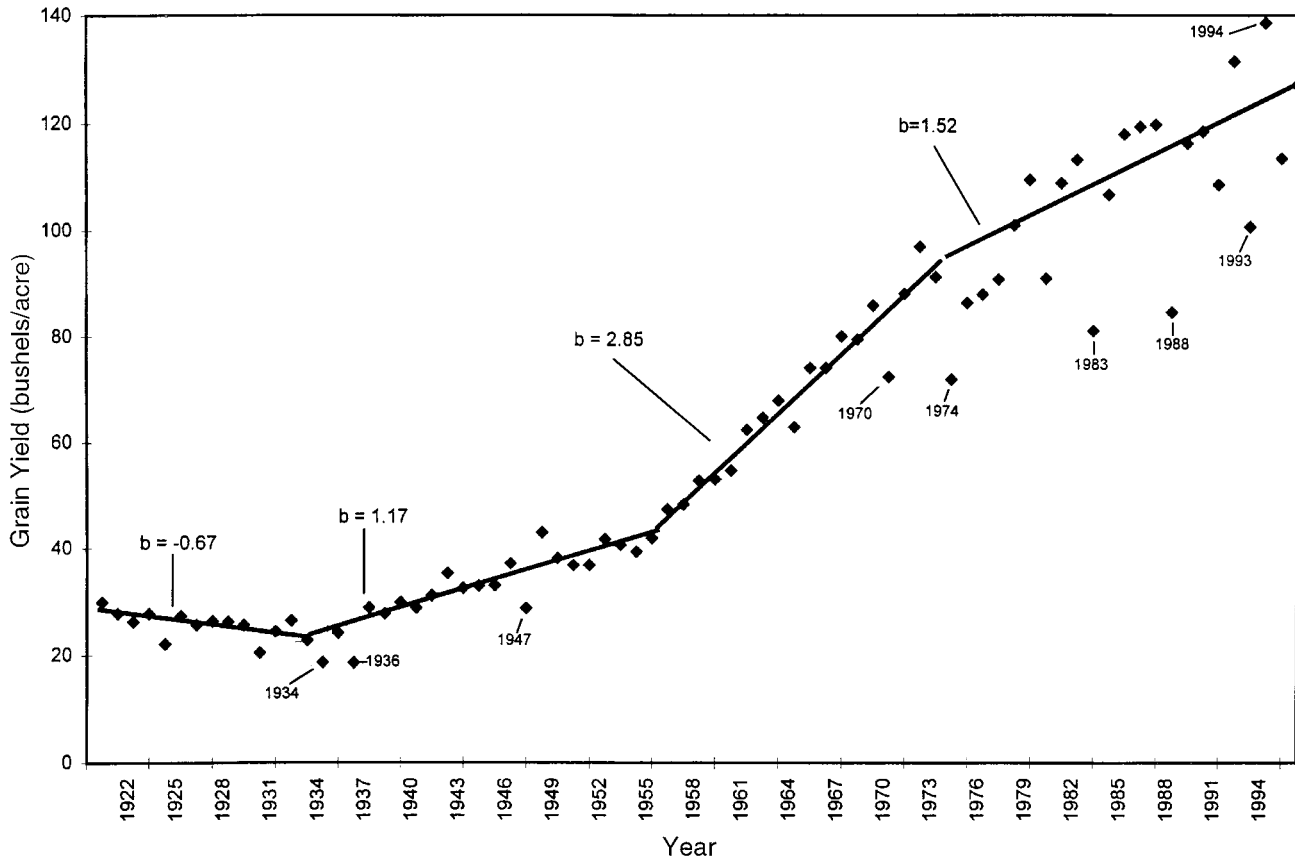


FIG. 1. Regression of U.S. maize grain yields on four groups of production years spanning 1920–1996. Points labeled in the plot were excluded from the analysis without altering the relative differences among the four groups of years. The regression coefficient for each regression line is indicated. (Reproduced with permission from O. S. Smith, United States Department of Agriculture Agricultural Statistics).

ity, and registration of crop varieties—activities that affect the conservation and use of gene pools. Historically, the majority of the information was based on phenotype, geographic origin, social history, and parentage. Collectively, those sources can provide useful information. However, protein-based genetic markers subsequently were developed, and they clearly illustrated the potential power and utility of molecular-based methods (6, 7). The nonmolecular methods shall remain important, but they have some serious weaknesses that have been, or could be, addressed by techniques derived from plant genomics.

Most nonmolecular methods suffer from a lack of supportive genetic information and weak discriminatory power. The genetic architecture and control of most phenotypic descriptors are unknown and are interactive with the environment and genetic background (6, 7). Estimators of parentage and coancestry are based on several simplifying assumptions that are incompatible with the forces of artificial or natural selection (6); information may be produced each generation regardless of the genetic events. DNA markers and sequencing produce quantitative views of genetic diversity and relationships more rapidly, comprehensively, and reproducibly than previous methods. Ultimately, they will create true gene banks or warehouses with functional inventories from largely uncharacterized gene pools.

The traditional boundaries between gene pools have been challenged by genetic transformation and genomic mapping. Comparative linkage analysis has revealed evidence of considerable conservation at the genotypic and phenotypic levels among groups of evolutionarily related but sexually isolated plant species (8, 9). Further analyses should identify levels of

allelic richness and important evolutionary histories that resulted in the unique biological repertoire of a species or gene. Such information and material could be used to endow another crop's gene pool with novel capabilities or depth that have survived the tests of time in another lineage (10).

Qualitative and other more informative views of gene pools may become available as genes and genomes are sequenced and functionally characterized as indicated by the analysis of disease resistance genes (11). Technologies for large scale surveys of gene expression patterns, chips (12) and traps (13), explore other dimensions of gene pools that have not been available for review. Instead of looking at individual genes, we will be compelled to review circuits of genes and their pathways to synthesize a meaningful comprehension of genetic variation. Collectively, these methods will provide the raw information and material for understanding the generation and manifestation of genetic variation in gene pools.

Plant Breeding Programs and Germplasm Use. Excluding mechanisms of reproductive isolation, several factors have limited the use of exotic germplasm (e.g., unadapted varieties, plant introductions, land races, and undomesticated relatives) for improving elite gene pools of annual, seed-propagated crops. One barrier may be the programs' success in satisfying myriad immediate demands with simple methods and presumably shallow gene pools. Varietal development typically follows a cyclical series of steps, parent selection, progeny development and evaluation, and selection of candidate varieties such that today's varieties are tomorrow's parents. Selection is based on numerous phenotypes, maturity, response to biotic and abiotic stress, and the quality and quantity of the harvested product; the net sum of one complete life cycle, all of the genes,

and their interactions are assessed. For example, malting barley varieties must satisfy as many as 22 quality traits (14) in addition to those related to adaptation and productivity.

Also, varietal development has become more competitive and costly; in the United States, development of one variety of maize or soybean requires 6–8 years, \$0.5–7.0 million (typically several million dollars for human and physical resources alone), and evaluation of 10^5 progeny. The lifetime of a variety is usually 3–6 years before it succumbs to the challenges of the production environment, demands of consumers, and competition from new varieties. Consequently, breeding programs devote most of their resources to manipulation of a core of elite germplasm known to provide genetic gain and to satisfy objectives for the near term (14, 15).

Another important barrier may be the perceived net value of exotic germplasm in the context of an elite gene pool and target environment. Exotic gene pools have been the ultimate sources of all varieties, but their recent contributions have been presumably genes with highly qualitative effects on defensive traits in most situations (14, 16). Meaningful and comprehensive assessments of exotic germplasm often are hindered severely by its poor adaptation to the target environment (e.g., photoperiod and temperature responses, seed or fruit retention). So, a few bad genes for adaptation obscure or preclude the evaluation of many others.

Once exotic germplasm has been introgressed into the elite genetic backgrounds, there may have been a tendency of researchers to attribute gains to new combinations of alleles from the elite gene pool (e.g., epistasis) rather than the effects of the exotic's. Before the advent of DNA markers, the ability to trace the parental source of favorable alleles was limited extremely. Genetic mapping studies with DNA markers have provided definitive proof that exotic germplasm contributes novel alleles with positive effects for several traits in ways that would not have been predicted on the basis of direct assessments of the exotic parent's phenotype (17). Therefore, the value of exotic gene pools may be severely underestimated. Perhaps greater awareness of the potentially pervasive benefits of exotic germplasm will lead to more effective use for a wider range of objectives.

Modification of Breeding Schemes. Of course, introgressing the exotic germplasm and realizing meaningful improvements in an efficient manner has been and will remain challenging: retaining the good genes, eliminating the bad ones, and overcoming restrictions to recombination between exotic and domestic homologous chromosomes (18). Graphical genotypes revealed by DNA markers have illustrated clearly that "linkage drag" (i.e., retention of unwanted and genetically linked germplasm) for the same chromosome may vary greatly among breeding programs, methods, and parents (19). Such linkage information creates options for truly novel, data-driven selection methods that reduce the number of generations of selection, maximize recovery of the desired parent's genome, and elucidate patterns of genetic recombination. Consequently, breeding programs have used marker information to backcross transgenes and other native genetic factors with highly qualitative effects (8). Some of these principles could be used to adapt exotic germplasm (20) and exploit it for the improvement of complex traits with quantitative inheritance patterns (16, 21).

Various strategies have been used to successfully infuse elite gene pools of several crop species with new and useful alleles from exotic sources (22). The value and novelty of the exotic alleles are rarely known *a priori* and may be best assessed in the context of the elite genetic background. However, many potentially valuable alleles may be eliminated randomly during the transfer to the elite background before evaluation. In some scenarios, DNA markers could be a useful adjunct to established methods of adapting and introgressing exotic germplasm.

Conversion programs adapt exotic germplasm by mating it with an elite parent and selecting for phenotypes suited to the target environment (e.g., appropriate temperature and photoperiod responses and plant stature). The selected progeny are mated again (i.e., backcrossed) to the exotic parent and the cycle is repeated for several (e.g., five) generations to achieve an acceptable degree of adaptation and to increase the probability that capture of exotic germplasm has been maximized (23). By genotyping the selected progeny before backcrossing, DNA markers could be used efficiently in conversion programs to ensure maximum recovery of the exotic parent's genome and to reduce the number of backcross generations (20). This modification would increase the proportion of the exotic genome exposed to new (and also exotic) production environments and genetic backgrounds, and it should permit programs to reallocate resources in accordance with the reduction of generations required for suitable conversion.

Other strategies use DNA markers to identify, transfer, and combine specific regions of exotic genomes in elite genetic backgrounds for improvement of complex traits with quantitative inheritance patterns (16, 21). The Advanced-Backcross QTL (quantitative trait locus) analysis (21), as demonstrated in tomato (24), extends an earlier method (25) by using markers to map QTLs after two or three generations of backcrossing and selection to eliminate unadapted phenotypes. Herein, favorable exotic alleles at QTLs are identified in the context of more uniform and elite genetic backgrounds, and linkage information may be used to combine favorable QTLs into one superior genotype, a candidate variety.

Marker-assisted introgression is not a panacea, but it will help. Like any breeding method, its utility is affected by many factors, such as reproductive biology of the crop (e.g., incompatibility, ease of controlled pollination, and number of seed produced per plant and per pollination), the genetic wealth of the elite gene pool, and the relative efficiency of other strategies. At this time, this approach also is limited by the ease with which marker data are collected and managed. DNA marker methods, except in the simplest scenarios, often are not suited for the speed, scale, and complex biology encountered in modern breeding programs (8, 26). Inexpensive, decentralized, and rapid diagnostic marker methods are needed.

Unresolved Issues and Needs. The imminent and current challenges of crop-based agriculture suggest that gains will be more difficult and expensive to achieve. Often, the best strategy for gain may rely more on making changes to the plant and less on changes to the agricultural system and supporting environment. Plants are complex biological systems, and to understand the systems, one must have as many of the pieces as possible. Genomics will provide many of the pieces and will enhance our abilities to appreciate, use, and enhance crop gene pools. However, the degree, merit, and beneficiaries of the enhancements are debatable.

Previously, gene pools have been free and open to humanity and often have been treated rather shabbily. They are fundamental natural resources equivalent to air, water, and soil and were developed through evolution, domestication, and civilization before the establishment of nations and corporations. They differ from electronics, automobiles, and pharmaceuticals in that they provide an absolutely essential ingredient for human life: food. Will they receive better treatment under a relatively new code of morals, ethics, and laws—the double-edged swords of plant intellectual property rights and raw capitalism?

The reductionist element of genomics seeks to define the function of every dimension in the genomes: a noble and worthy goal. However, a great moment of unity in science and agriculture easily can accelerate some dangerous trends and precedents that conceivably culminate by partitioning and protecting all of the isolated bits of "useful" biology. Nobody invented the gene pools or their fundamental utilities; yet, we

are willing to patent and perhaps sequester parts or all of them. Furthermore, data have not been exchanged openly or deposited rapidly in some genome projects. Parallel trends have emerged for exchange of crop germplasm. If crop plant genome projects are conducted similarly, then we might expect unnecessary duplication of effort, a wasteful diversion of resources from pathetically weak support for basic and applied plant sciences, and further destabilization of the delicate equilibria of crop-based agriculture and the supporting environment. A wise balance is needed soon.

The author thanks O. S. Smith for providing Fig. 1 and K. J. Frey for reviewing the manuscript. Journal paper no. J-17516 of the Iowa Agriculture And Home Economics Experiment Station, Project 3134.

1. Brown, L. R. (1996) *Tough Choices: Facing the Challenge of Food Scarcity* (Norton, New York).
2. Huffman, W. E. & Evenson, R. E. (1993) *Science for Agriculture: A Long Term Perspective* (Iowa State Univ. Press, Ames, IA).
3. Duvick, D. N. (1984) in *Gene Manipulation in Plant Improvement*, ed. Gustafson, J. P. (Plenum, New York), pp. 17–30.
4. Horsfall, T. G., chair (1972) *Genetic Vulnerability of Major Crops* (National Academy of Sciences, Washington, D.C.).
5. Duvick, D. N. (1992) *Maydica* **37**, 69–79.
6. Smith, J. S. C. & Smith, O. S. (1992) *Adv. Agron.* **47**, 85–140.
7. Bretting, P. K. & Widrlechner, M. P. (1995) in *Plant Breeding Reviews*, ed. Janick, J. (Wiley, New York), Vol. 13, pp. 11–86.
8. Lee, M. (1995) *Adv. Agron.* **55**, 265–344.
9. Paterson, A. H., Lin, Y.-R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R. M., Liu, S.-C., Stansel, J. W. & Irvine, J. E. (1995) *Science* **269**, 1714–1718.
10. Sorrells, M. E. & Wilson, W. A. (1997) *Crop Sci.* **37**, 691–697.
11. Staskawicz, B. J., Ausubel, F. W., Baker, B. J., Ellis, J. G. & Jones, J. D. G. (1996) *Science* **268**, 661–667.
12. Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. & Fodor, S. P. A. (1996) *Science* **274**, 610–614.
13. Sundaresan, V. (1996) *Trends Plant Sci.* **1**, 184–190.
14. Rasmusson, D. C. & Phillips, R. L. (1997) *Crop Sci.* **37**, 303–310.
15. Hallauer, A. R. (1990) *Maydica* **35**, 1–16.
16. Edwards, M. D. (1992) *Field Crops Res.* **29**, 241–260.
17. de Vicente, M. C. & Tanksley, S. D. (1993) *Genetics* **134**, 585–596.
18. Williams, C. G., Goodman, M. M. & Stuber, C. W. (1995) *Genetics* **141**, 1573–1581.
19. Young, N. D. & Tanksley, S. D. (1989) *Theor. Appl. Genet.* **77**, 353–359.
20. Pereira, M. G. & Lee, M. (1995) *Theor. Appl. Genet.* **90**, 380–388.
21. Tanksley, S. D. & Nelson, J. C. (1996) *Theor. Appl. Genet.* **92**, 191–203.
22. Shands, H. L. & Wiesner, L. E., eds. (1991) *Use of Plant Introductions in Cultivar Development, Part 1*, eds. Shands, H. L. & Wiesner, L. E. (Crop Science Society of America, Madison, WI).
23. Duncan, R. R., Bramel-Cox, P. J. & Miller, F. R. (1991) in *Use of Plant Introductions in Cultivar Development, Part 1*, eds. Shands, H. L. & Wiesner, L. E. (Crop Science Society of America, Madison, WI), pp. 69–102.
24. Tanksley, S. D., Grandillo, S., Fulton, T. M., Zamir, D., Eshed, Y., Petiard, V., Lopez, J. & Beck-Bunn, T. (1996) *Theor. Appl. Genet.* **92**, 213–224.
25. Bliss, F. A. (1981) *Hort. Sci.* **16**, 129–132.
26. Rafalski, J. A. & Tingey, S. V. (1993) *Trends Genet.* **9**, 275–280.